



Chen, X., Chen, B. and Kankanhalli, M. (2017) MM2RTB: Bringing Multimedia Metrics to Real-Time Bidding. In: 10th International Workshop on Data Mining for Online Advertising (ADKDD '17), Halifax, NS, Canada, 14 Aug 2017, ISBN 9781450351942.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

© ACM 2017. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in 10th International Workshop on Data Mining for Online Advertising (ADKDD '17), Halifax, NS, Canada, 14 Aug 2017, ISBN 9781450351942, <https://doi.org/10.1145/3124749.3124757>.

<http://eprints.gla.ac.uk/159031/>

Deposited on: 14 March 2018

# MM2RTB: Bringing Multimedia Metrics to Real-Time Bidding

Xiang Chen  
National University of Singapore  
chxiang@comp.nus.edu.sg

Bowei Chen  
University of Lincoln  
bchen@lincoln.ac.uk

Mohan Kankanhalli  
National University of Singapore  
mohan@comp.nus.edu.sg

## ABSTRACT

In display advertising, users' online ad experiences are important for the advertising effectiveness. However, users have not been well accommodated in real-time bidding (RTB). This further influences their site visits and perception of the displayed banner ads. In this paper, we propose a novel computational framework which brings multimedia metrics, like the *contextual relevance*, the *visual saliency* and the *ad memorability* into RTB to improve the users' ad experiences as well as maintain the benefits of the publisher and the advertiser. We aim at developing a vigorous ecosystem by optimizing the trade-offs among all stakeholders. The framework considers the scenario of a webpage with multiple ad slots. Our experimental results show that the benefits of the advertiser and the user can be significantly improved if the publisher would slightly sacrifice his short-term revenue. The improved benefits will increase the advertising requests (demand) and the site visits (supply), which can further boost the publisher's revenue in the long run.

## KEYWORDS

Display advertising; real-time bidding; multimedia metrics

### ACM Reference format:

Xiang Chen, Bowei Chen, and Mohan Kankanhalli. 2017. MM2RTB: Bringing Multimedia Metrics to Real-Time Bidding. In *Proceedings of ADKDD'17, Halifax, NS, Canada, August 14, 2017*, 6 pages. DOI: 10.1145/3124749.3124757

## 1 INTRODUCTION

The fundamental issue of computational advertising is to select a proper ad from a set of ad candidates. Introduced in 2009, RTB has become the norm of selling online users' page views (also called *impressions*) in display advertising, where the ad with the highest bid wins the auction for the impression. Obviously, the existing RTB system is biased towards the publisher. However, the effectiveness of displayed ads remains debatable, in terms of the benefits of the advertiser and the user. According to hubspot [8], a recent survey shows only 2.8% of participants think banner ads are relevant to what they read online and another 33% even think many ads completely intolerable. In the year 2015, ad blocking has grown by 41% globally. On one hand, the user gets annoyed by the improper ads. On the other hand, the advertiser gets harmed by ineffective ad delivery. Moreover, for those webpages with

multiple ad-slots, the ad slots are sold separately through individual auctions [1]. As a result, two ads about the same product but belong to different advertisers are likely to be displayed within the same webpage. The competitive ads have a negative effect on user's brand conception [3].

Motivated by the above observations, we argue that the context matters in display advertising. Literature in marketing and consumer psychology has already shown that the contextual relevance between the content of host webpage and the ads makes a large difference in their clickability [5], and it also has a leading effect on the user's online experience [14]. Through a large-scale live test, combining contextual relevance and targeting strategies is able to increase the ad CTR [12]. To increase user's engagement towards the displayed ads, we introduce the *contextual relevance*. Through a series of eye-tracking experiments, recent research found that users tend to avoid ads in web search and surfing [2] and they intentionally avoid looking at such ads even when they are designed to be attention-grabbing [17]. This is also known as *ad overlook*. Moreover, Owen et al. [16] explored the relationship between ads location and the degree of blindness – the phenomenon of website users actively ignore web banner ads – and found that users tend to ignore ads located on the bottom and right area. Intuitively, any ad that fails to capture the user's attention will be ineffective in delivering information. To ensure that the user will notice the displayed ad, we introduce the *visual saliency*. Image memorability has been shown to be a stable and intrinsic property of images that is shared across different viewers [11]. Thus, the memorable ads will be easily recalled by the user. To enhance the user's brand conception towards the displayed ad, we introduced the *ad memorability*.

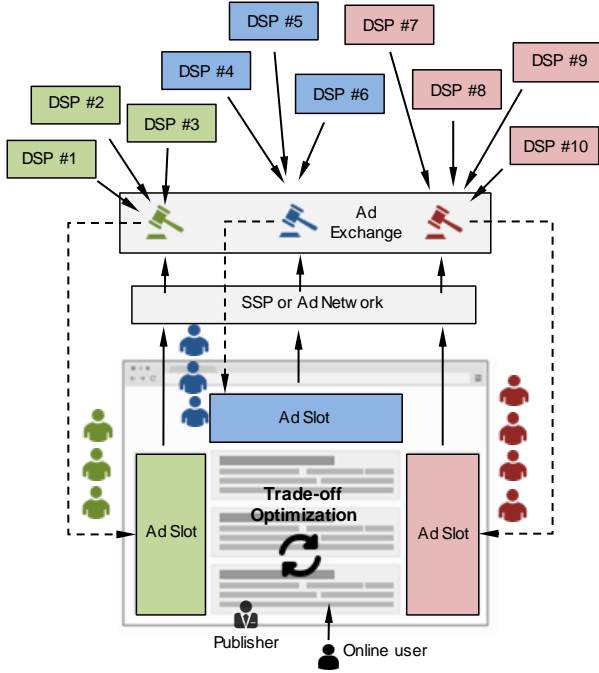
Fig. 1 illustrates how multimedia metrics are incorporated into RTB. When an online user visits a publisher's website, the publisher's web server loads up a webpage and sends an HTML code to the user's browser so that the latter will know where to get the content and how to format it. In the meantime, the publisher's webpage information, such as the publisher ID, the site ID, and the dimensions of ad slots, together with the user's cookie ID, will be passed to and be accessed by an RTB-enabled supply-side platform (SSP) or ad network. Each ad slot is considered to generate an impression and it will be treated independently in the selling. For an ad slot, an auction is started within ad exchange by requesting bids from demand sources such as the demand-side platforms (DSPs) or individual advertisers, the winning advertiser will be able to have his ad displayed to the user.<sup>1</sup> The whole RTB process includes user identification, auction and ad display, usually be finished in 10 to 100 milliseconds [19]. The displayed ads on a single webpage come from individual auctions and the selection criteria is mainly based on bids. The visual effects, particularly, their interactive effects are not well considered. This will affect the user's experience as well as the advertising effectiveness for each advertiser, which further

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ADKDD'17, Halifax, NS, Canada

© 2017 ACM. 978-1-4503-5194-2/17/08...\$15.00  
DOI: 10.1145/3124749.3124757

<sup>1</sup>Individual advertisers typically do not directly participate in advertising auctions but entrust some demand-side companies to bid on their behalf.



**Figure 1: Schematic view of trade-offs optimization for RTB.**

affect the publisher's long-term revenue. Therefore, we propose a trade-offs optimization framework in this paper.

In this study, we focus on improving the effectiveness of display advertising with considering all stakeholders' benefits. We consider multimedia metrics to measure some stakeholders' benefits, such as the ad memorability, the contextual relevance, and the visual saliency. To the best of our knowledge, this is one of the few studies that combines multimedia techniques and auction theory together. Different from our previous work [6], which focuses on single-slot display advertising, we extend the idea under multi-slot scenario.

## 2 PRELIMINARIES

The key concepts and notations are briefly explained as below:

**Publisher** – a company or individual who owns a webpage which has slots to host online ads. A webpage is denoted by  $w$ .

**Ad slot** – a rectangular area within a webpage where a banner ad can be displayed to online users. In this paper, we consider the case that a webpage contains multiple ad slots. A slot on webpage  $w$  is denoted by  $s \in S_w$ , where  $S_w$  is the set of slots.

**Advertiser** – a company or individual who wants to display his banner ad on the publisher's webpage. The notation  $a_{s,w,l}$  represents advertiser  $l$  bids for slot  $s$  on webpage  $w$ .

**Company** – a company is defined based on the URL domain of the ad landing page. If two different ads point to a same URL domain, they are recognized as one company's ads.

**Topic** – a cluster in which ads are about similar products or services. We use the ad landing page texts to represent the ad information, and employ the term frequency-inverse document frequency (TF-IDF) and K-Means methods [13] to cluster all ads into 24 topics<sup>2</sup>.

<sup>2</sup>According to the IAB content taxonomy [9], online contents can be broadly grouped into 24 topics.

**Competitive advertiser** – an advertiser is said to be a competitive advertiser to the other if both advertisers have at least one ad being clustered into a same topic group. For example, if iPhone 7 and Galaxy S8 are clustered into the smartphone topic, Apple and Samsung are competitors. However, even though iPhone 6S and iPhone 7 join RTB as different bidders, they are not competitive advertisers because both come from Apple.

## 3 TRADE-OFFS OPTIMIZATION

Suppose that a webpage  $w$  has a set of slots  $S_w$  to host banner ads. When an online user visits it,  $|S_w|$  impressions are created and are auctioned off separately. The model optimizes the trade-offs among multiple stakeholders over multiple slots by the following steps:

**Step 1** – Create a matrix  $\Omega$  that contains all advertisers who join RTB campaigns for impressions from any slot(s) of webpage  $w$ . As shown in Fig. 1, the example webpage contains three ad slots. For this online user's visit, there are: three advertisers bid for the first slot (in blue color); three advertisers bid for the second slot (in green color); and three advertisers bid for the third slot (in red color). Therefore,  $\Omega$  can be expressed as

$$\Omega = \begin{pmatrix} a_{1,w,1} & a_{2,w,1} & a_{3,w,1} \\ a_{1,w,1} & a_{2,w,1} & a_{3,w,2} \\ a_{1,w,1} & a_{2,w,1} & a_{3,w,3} \\ a_{1,w,1} & a_{2,w,1} & a_{3,w,4} \\ a_{1,w,1} & a_{2,w,2} & a_{3,w,1} \\ a_{1,w,1} & a_{2,w,2} & a_{3,w,2} \\ & \dots & \\ a_{1,w,3} & a_{2,w,3} & a_{3,w,4} \end{pmatrix},$$

where  $a_{1,w,3}$  represents advertiser 3 who bids for the first slot on webpage  $w$ . The column of  $\Omega$  represents a specific slot and its row represents a specific combination of candidate advertisers, so the size of  $\Omega$  is  $z \times |S_w|$ , where  $z = \prod_{s \in S_w} (|\mathcal{A}_{s,w}|)$ , and  $\mathcal{A}_{s,w}$  is the set of advertisers who join the RTB campaign for slot  $s \in S_w$ .

**Step 2** – Create a subset matrix  $A$  from  $\Omega$  by removing the rows which contain competitive advertisers. The size of  $A$  is  $q \times |S_w|$ , where  $0 \leq q \leq z$ .

**Step 3** – Compute the rank score for each combination of candidate advertisers in  $A$ . For  $i = 1, \dots, q$ , the rank score  $r_i$  is

$$r_i = \mathbf{y}^{*\top} \mathbf{x}_i, \quad (1)$$

where  $\mathbf{x}_i$  is the vector of calculated values of metric variables for candidate advertisers on the  $i$ th row of  $A$ , and  $\mathbf{y}^*$  is the vector of their optimal weights. More details about  $\mathbf{x}_i$  and  $\mathbf{y}^*$  are discussed in Sections 4-5, respectively.

**Step 4** – The optimal selection of advertisers are  $A(i^*, :)$ , where

$$i^* = \arg \max_i \{r_i\}_{i=1}^q. \quad (2)$$

## 4 CALCULATION OF METRIC VARIABLES

Six metric variables are considered in the trade-offs optimization, including the publisher's revenue, the advertisers' utility, the ads' memorability, the CTR, the contextual relevance, and the ads' saliency. Different to [6], as multiple slots are considered in this study, each variable's value is the sum of the corresponding metric values of selected advertisers, and each slot has the same weights in the variable value calculation. Algorithm 1 illustrates the calculation of metric variables for a given vector of candidate advertisers

$A(i, :)$ . Note that, all the six metrics are normalized before proceeded to the objective function.

As discussed earlier, impressions are auctioned off separately. If  $a \in A(i, :)$  and she bids for slot  $s$ , we can employ the method discussed in [6] by creating several pseudo slots and then calculate his (potential) payment by a GSP auction model. An advertiser's short-term benefit can be measured by his utility, which is defined as the difference between his value and payment. An advertiser's long-term benefit can be measured by the ad's memorability [15] – it shows how likely the user will remember the advertiser's ad in a few weeks or months. Same as [6], we employ the MemNet [10] model to predict the visual memorability of the ad image. The rest three metrics represent the user's benefit. The ad CTR is defined as the number of clicks on the advertiser's ad divided by the number of displays, whose value is usually given by data or can be estimated from historical advertising records. It is an ad quality metric – a high CTR means that the advertiser's ad is attractive or more relevant to the user's needs. The contextual relevance measures if the ad content is more or less relevant to its hosting webpage content. Note that, the textual information of the webpage is represented by the combination of webpage title, keywords, description and main content. Different to [6], we use the TF-IDF [13] here to measure the similarity of textual contents between the ad and the webpage. This is because [6] uses webpage title, keywords and description to construct the textual information of the webpage, where the adopted method Takelab [18] is good at measuring the similarity between short text snippets. The ad saliency metric measures whether the ad image can be easily spotted within its hosting webpage. We use the minimum barrier salient (MBS) object detection method [20] to calculate the saliency of the ad image. For each pair of webpage and ad candidate, we embed the ad into the webpage and use the MBS method to calculate the saliency map. We take the mean value of each pixel within the ad area and view this value as the saliency score of the ad candidate.

## 5 DETERMINATION OF OPTIMAL WEIGHTS

Suppose that  $n$  site visits have been observed, the optimal weights  $\mathbf{y}^*$  can be obtained by solving the following optimization problem:

$$\mathbf{y}^* = \arg \max_{\mathbf{y}} \sum_{j=1}^n \mathbf{y}^\top \mathbf{x}_{i^*}^{\{j\}}, \quad (3)$$

$$\text{s.t. } 0 \leq \gamma_k \leq 1, k = 1, \dots, 6, \quad (4)$$

$$\mathbf{y}^\top \mathbf{1} = 1, \quad (5)$$

$$|\xi_1| \leq |\theta_1|, \theta_1 \leq 0, \quad (6)$$

$$\xi_k \geq \theta_k, \theta_k \geq 0, k = 2, \dots, 6, \quad (7)$$

where  $\xi_k$  is defined by

$$\xi_k = \frac{\sum_{j=1}^n (\mathbf{x}_{i^*}^{\{j\}}(k) - \mathbf{x}_{i^-}^{\{j\}}(k))}{\sum_{j=1}^n \mathbf{x}_{i^-}^{\{j\}}(k)}, \quad k = 1, \dots, 6. \quad (8)$$

The terms  $i^*$  and  $i^-$  are the indexes of optimal solution and ground truth (i.e., the selected ads using the existing RTB system) for candidate matrix  $A^{\{j\}}$  for the  $j$ th observation in the training data, and  $\theta_k$  is the threshold value for specifying the bound of changes in variable  $k$ . The optimal weights maximize the sum of rank scores of the select advertisers from all auctions in the training

### Algorithm 1 Calculation of metric variables for a given $A(i, :)$ .

```

1:  $\mathbf{x}_i(1) \leftarrow \sum_{a \in A(i, :)} \text{Payment}_a$  ▷ GSP auction [6]
2:  $\mathbf{x}_i(2) \leftarrow \sum_{a \in A(i, :)} (\text{Value}_a - \text{Payment}_a)$ 
3:  $\mathbf{x}_i(3) \leftarrow \sum_{a \in A(i, :)} \text{MemScore}_a$  ▷ MemNet [10]
4:  $\mathbf{x}_i(4) \leftarrow \sum_{a \in A(i, :)} \text{CTR}_a$ 
5:  $\mathbf{x}_i(5) \leftarrow \sum_{a \in A(i, :)} \text{RevScore}_a$  ▷ TF-IDF [13]
6:  $\mathbf{x}_i(6) \leftarrow \sum_{a \in A(i, :)} \text{SaliencyScore}_a$  ▷ MBS [20]
7: return  $\mathbf{x}_i$ 

```

**Table 1: Multimedia datasets. The crawler type (I) collects as many webpages as possible from a seed URL and the crawler type (II) repeatedly accesses to a set of particular webpages.**

Website	Yahoo	Yahoo	MSN	MSN
Crawler type	I	II	I	II
From	20 Jan 2017	20 Jan 2017	20 Jan 2017	20 Jan 2017
To	30 Jan 2017	30 Jan 2017	30 Jan 2017	30 Jan 2017
Location	Singapore	Singapore	Singapore	Singapore
# of webpages				
with 1 slot	1,481	1,909	798	686
with 2 slots	1,978	3510	1,519	3,689
with 3 slots	1,173	4329	3,633	146
with $\geq 4$ slots	1,599	2,468	41	241
# of total impressions	15,836	31,951	14,899	9,466
# of unique advertisers	692	631	160	163
# of total companies	475	431	96	99

set. Eqs. (4)-(5) ensure each variable has an impact and its impact has an upper bound. Eqs. (6)-(7) further specify the maximum decrease of the publisher's revenue and the minimum increases for other variables.

## 6 EXPERIMENTS

This section presents our datasets, empirical findings, experimental settings, and overall results of trade-offs optimization.

### 6.1 Datasets

We use two types of datasets to simulate a real-world advertising system: the multimedia dataset represents the interaction between the publisher and the user, and the auction dataset shows the interplay between the publisher and the advertiser.

We have collected data from Yahoo and MSN over the period from 20 January to 30 January in 2017 to construct our multimedia datasets. All the multimedia datasets were collected in Singapore. For each website, we designed two data crawlers, crawler type (I) started from a given seed URL, and used breadth-first-search to collect as many different webpages as possible. To ensure the diversity of webpages and banner ads, we started from the home page of each website, which contains multiple categories of contents.<sup>3</sup> Crawler type (II) repeatedly accessed to a set of particular webpages at a frequency of every 5 minutes. In our experiment, this webpage set was made up of the homepage of the sub categories, such as Yahoo news, Yahoo finance, Yahoo sports and so on, which were the most frequently viewed webpages by the users. These two types of data crawler were designed for different purpose: crawler type (I) indicated how the ad changed with different webpage contents, while crawler type (II) represented how the ad changed with different timestamps. Since the ads on both websites are dynamically embedded, we used Python, Selenium and

<sup>3</sup> Yahoo: <https://sg.yahoo.com>

MSN: <http://www.msn.com/en-sg/>

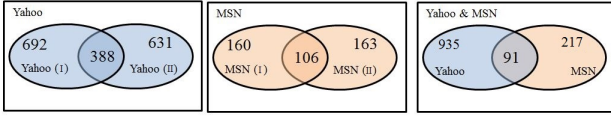


Figure 2: The intersection of ad sets within each website and across different websites

Table 2: Statistic about webpage with different scenarios

Ad dataset	Yahoo I	Yahoo II	MSN I	MSN II
# of total webpages	4,750	10,307	5,193	4,076
# of webpages with scenario 1)	786	2,169	3,669	1,480
# of webpages with scenario 2)	155	924	572	238
# of webpages with scenario 3)	412	836	158	82

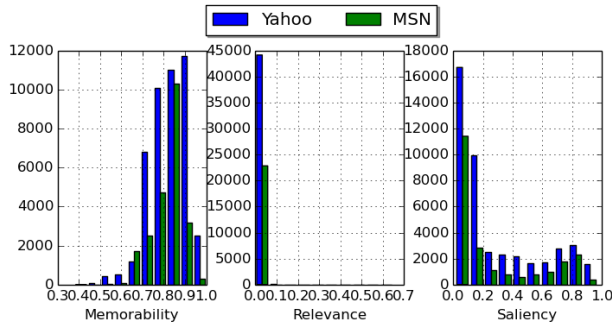


Figure 3: The distribution of multimedia metrics on the two websites, where X-axis represents the value of variables, and Y-axis represents the number of webpage-ad pairs.

Chromedriver to collect the webpage and ads that displayed to the user as shown in real-world. Note that, in collecting the data, the browser was set in the privacy mode, which disabled browsing history, web cache and data storage in cookies so that the collected banner ads were not affected by the previous page views. In each dataset, we extracted the ads from their webpages to create a set of banner ads and a set of webpages with blank ad slots. For each webpage, the collected data includes title, keywords, description, whole webpage snapshot, ad image. We also crawled title, keywords and description from the ad landing page (i.e., the delivered webpage if an ad is clicked by user). Note that we did not consider animation ads in our banner ads. According to the definition in section 2, we were able to figure out the advertiser set, company set and competitive advertiser-pair set. Our multimedia datasets have been released publicly and more information can be found at: <https://github.com/boweichen/MultiSlotRTBMultimediaDataset>.

## 6.2 Empirical Findings

Table 1 describes the data pre-processing results on the four multimedia datasets. As we can see, the total number of ad impressions of each website under each crawler type is 15,836, 31,951, 14,899, 9,466, while the number of unique advertisers is 692, 631, 160, 163, respectively. Although there are a large number of impressions in the ad network, only a few ads are displayed. We find that some ads re-appear from time to time. The repetitive display strategy reinforces users' memory for branding but it is also a source of intrusiveness into users' online experience [4]. Another interesting finding is illustrated in Fig. 2. As can be seen, there are 388 advertisers' ads appearing in both Yahoo datasets out of all the 935

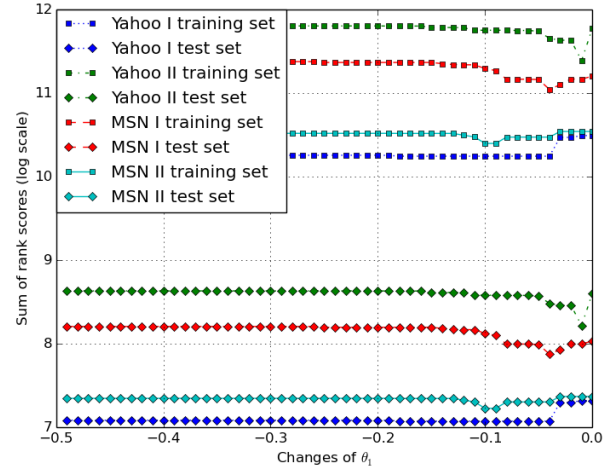


Figure 4: The effect of  $\theta_1$  on the sum of total rank scores of the selected advertisers in the auctions. Note that we only show  $\theta_1$  from 0.0 to -0.50, since the sum of rank score will not change when  $\theta_1$  is greater than a particular value.

advertisers, and 91 out of 217 in two MSN datasets. The intersection within each website indicates that: the content of the host webpages affects the displayed ads and same ads only appeared in the specific webpages. And also, there are 91 advertisers' ads appearing on both Yahoo and MSN, which shows the existence of interactions between different DSPs and websites.

Since user's memory towards displayed ads varies in various contexts, the repetitive ads can enhance the brand perception while competitive ads will result in the opposite effect [3]. Given that each ad slot is auctioned separately, the following three scenarios are possible in a multi-slot webpage: 1) two ads with the same landing page are displayed; 2) two ads with different landing page but belonging to the same company are displayed; 3) two competitive ads are displayed. Note that there exists overlaps among the above three scenarios. For example, there are four ads  $ad_1, ad_2, ad_3, ad_4$  displayed in the webpage, where  $ad_1, ad_2$  are promotions to buy the latest Apple Iphone 7 in Apple store,  $ad_3$  is a promotion to buy a Iphone 6, while  $ad_4$  is about how to buy Samsung Galaxy S7. All the above three scenarios occur. Table. 2 summarizes the statistic of the three scenarios in our four multimedia datasets. When comparing the number of webpages in scenario 1) and scenario 2), we can find that: when two ads from a same company are displayed within the same webpage, the probability that they are the same ad is higher than that of different ads. This is because the advertiser won two independent auctions, which belong to the same webpage. Moreover, when two ads share the same topic are displayed within a webpage, the probability of being competitive ranges from 21.3% to 30.5% in Yahoo website, and from 3.59% to 4.55% in MSN website. Considering the extremely large number of daily page views, the number of webpages with competitive ads is still impressive. Since user's brand conception towards displayed ad will be poor in competitive context, the effectiveness of advertising of the whole system will be affected.

Fig. 3 shows the distribution of multimedia metrics, namely, the ad memorability, webpage-ad relevance and ad visual saliency in Yahoo and MSN. As we can see, in both two websites, the memorability score of most ads is above 0.70, indicating that these ads are



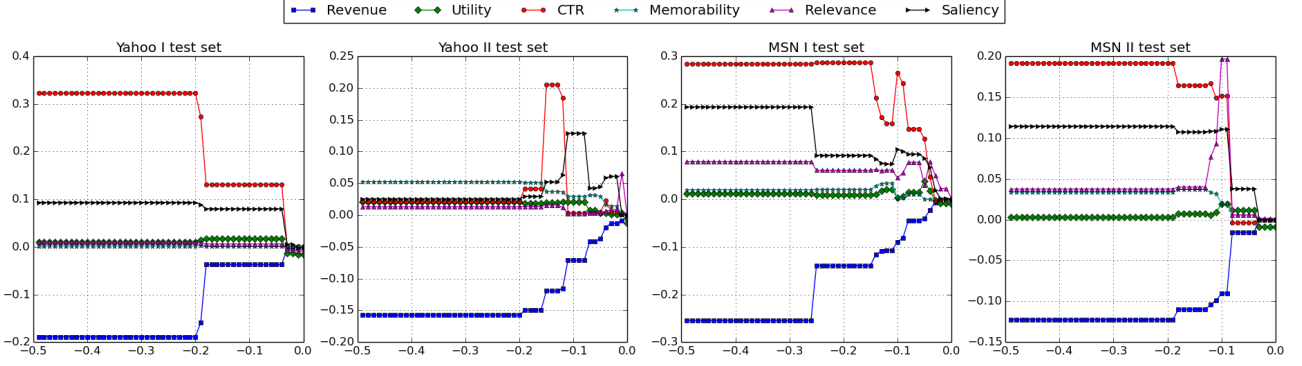


Figure 5: The effect of  $\theta_1$  on the changes of variables, where X-axis represents the change of  $\theta_1$  and Y-axis represents the change of corresponding variables. Note that: 1) we only show  $\theta_1$  from 0.0 to -0.50, since the variables will not change when  $\theta_1$  is greater than a particular value; 2) we do not show the changes of variables in the training set since the results of the training sets are similar to that of the test sets.

well designed to be memorable by the advertisers. While contextual relevance score of the majority ads is pretty low, which is consistent with the finding from marketing data that only 2.8% users thought ads on website were relevant [8]. In addition, most displayed ads are not salient, which can be easily overlooked by the user. The irrelevant and non-salient ads will result in less user engagement. In this regard, our proposed model incorporating multimedia metrics can help improve the effectiveness of existing advertising systems.

### 6.3 Experimental Settings

To validate the proposed framework, both bidding information (e.g., bid price, ad CTR) and multimedia information (e.g., text for contextual relevance matching, ad image for visual saliency and image memorability) for each ad are needed. As each of our datasets only provides partial information, we do random sampling to connect the auction dataset with the multimedia datasets. For each ad slot within a multi-slot webpage, we simulate the corresponding auction following the procedures described in our previous work [6]: for a given webpage, the original ad (i.e., the one displayed in the webpage when we collected the data) is treated as the ground truth, and is allocated with the highest bid price. We then sample the rest candidate ads and randomly match them with bid prices. It should be noted that slots have different shapes and the ads with similar shapes can be selected as candidate ads. We then apply the framework in Section 3 to select the proper combination of ads.

Every millisecond matters in RTB. Using the TF-IDF technique [13] to measure the contextual relevance has been demonstrated to be efficient; using MemNet [10] to obtain the ad image memorability score can be conducted off-line; however, using MBS [20] to calculate the visual saliency can be relatively time consuming. The processing speed of MBS is 80 frames per second on a machine with 3.2GHz×2CPU and 12GB RAM. Given a webpage with 2 ad-slot, and both the two slots have 10 bidders. According to our framework, we need to calculate the saliency map for 100 images. Thus, the time consumption becomes intolerable. In our experiment, we find that: in terms of saliency, the interaction among the ad slots is small. This is because the relative locations of two ad slots is relatively far and the saliency of each slot is mainly affected by the surroundings. In this regard, we only have to calculate the saliency

for 20 images in the previous example. We can further sacrifice the performance in speeding up the saliency calculating process, such as using super-pixel segmentation [7].

### 6.4 Results

We conduct 10-fold cross validation on all the four datasets to demonstrate the effectiveness of our proposed framework. We first obtain the optimal weights from the training set using the algorithm described in section 5, then apply the optimal weights to select the most suitable ads in the test set using the optimization model described in section 3. Note that the optimal weights change with the threshold value  $\theta_k$ ,  $k = 1 \dots 6$ , which indicate publisher's preferences towards the corresponding variables. As a monetization industry, revenue is always the primary concern. For simplicity but without lose of generality, we set  $\theta_1$  to be a non-positive value and  $\theta_k = 0$ ,  $k = 2 \dots 6$  in our experiment. So that we will figure it out that: under the premise of the advertiser's satisfaction and the user's online experience, how much  $\theta_1$  will change the performance of the whole system as well as the performance of other variables.

Fig. 4 shows the effect of  $\theta_1$  on the performance of the advertising system. As can be seen, in both training set and test set under all datasets, when  $\theta_1$  decreases, the sum of rank scores of the selected ads will not change at the very beginning, decrease slightly afterwards, increase gradually later and remain stable finally. The convexity changing pattern of the sum of rank scores is such because: when  $\theta_1$  is small, our weights determination model is not able to find a solution to the problem. In other words, the publishers are too stingy to lose revenue, and too greedy to achieve improvement on the other variables. In this case, we will use the traditional auction mechanism to select the ad with highest bidding price in each auction. When  $\theta_1$  further decreases, the solution space increases and we will select the solution with the highest sum of rank scores. Note that, the ground truth may not be a solution to our model since competitive ads may occur. When  $\theta_1$  decreases below a threshold value, the optimal solution will not change, which results in the stability of the sum of rank scores. It should be noted that within the same dataset, the sum of rank scores in the training set is roughly 9 times that of the test set, since the size of training set is 9 times of the size of test set in a 10-fold cross validation.

**Table 3: Summary of optimal trade-offs among stakeholders in the Yahoo (I) dataset where  $\theta_1 = -0.05$ . Note that  $\gamma_i, i = 1, \dots, 6$  represent the optimal weights obtained from training set, and  $\xi_i, i = 1, \dots, 6$  represent the changes of corresponding variables.**

Fold	Optimal weight						Training set						Test set					
	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$\gamma_5$	$\gamma_6$	$\xi_1$	$\xi_2$	$\xi_3$	$\xi_4$	$\xi_5$	$\xi_6$	$\xi_1$	$\xi_2$	$\xi_3$	$\xi_4$	$\xi_5$	$\xi_6$
1	0.40	0.35	0.00	0.05	0.15	0.05	-3.8%	1.6%	0.0%	14.6%	0.4%	8.2%	-3.5%	0.8%	0.2%	14.4%	1.0%	6.1%
2	0.40	0.35	0.00	0.05	0.15	0.05	-3.7%	1.4%	0.0%	14.5%	0.5%	7.8%	-4.8%	2.3%	0.0%	15.0%	0.0%	9.9%
3	0.40	0.35	0.00	0.05	0.15	0.05	-3.8%	1.5%	0.0%	14.6%	0.3%	8.1%	-3.9%	1.8%	0.0%	14.2%	1.6%	7.0%
4	0.40	0.35	0.00	0.05	0.15	0.05	-3.9%	1.5%	0.0%	14.7%	0.7%	8.0%	-3.4%	2.1%	0.3%	13.5%	-2.1%	7.7%
5	0.40	0.35	0.00	0.05	0.15	0.05	-3.7%	1.3%	0.0%	14.4%	0.4%	7.9%	-4.6%	3.8%	0.2%	15.8%	1.3%	9.1%
6	0.50	0.40	0.00	0.00	0.05	0.05	-1.9%	1.5%	0.0%	0.0%	0.7%	7.9%	-1.7%	1.2%	-0.3%	0.0%	3.4%	6.9%
7	0.40	0.35	0.00	0.05	0.15	0.05	-3.9%	1.8%	0.0%	14.5%	0.7%	7.9%	-3.1%	-0.2%	0.0%	15.0%	-1.5%	8.6%
8	0.40	0.35	0.00	0.05	0.15	0.05	-3.8%	1.5%	0.0%	14.7%	0.5%	8.0%	-4.1%	2.2%	0.4%	13.0%	-0.3%	8.3%
9	0.40	0.35	0.00	0.05	0.15	0.05	-3.8%	1.5%	0.0%	14.4%	0.4%	8.2%	-3.5%	2.1%	0.0%	15.7%	0.8%	6.1%
10	0.40	0.35	0.00	0.05	0.15	0.05	-3.7%	1.6%	0.0%	14.6%	0.3%	7.8%	-4.3%	1.0%	0.1%	14.1%	1.8%	9.6%
Mean	-	-	-	-	-	-	-3.6%	1.5%	0.0%	13.1%	0.5%	8.0%	-3.7%	1.7%	0.1%	13.1%	0.6%	7.9%
Std.	-	-	-	-	-	-	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000

Fig. 5 describes the effect of  $\theta_1$  on the change of variables. When  $\theta_1$  is close to 0, all the variables remains unchanged, which is consistent with the results in Fig. 4. When  $\theta_1$  further decreases, the revenue follows a monotone decreasing pattern and the other variables increase to various degrees. We observe that when publishers specify the value of  $\theta_1$ , our optimal weights may result in much less revenue loss. For example, in Yahoo I dataset, when we set  $\theta_1$  as -0.15, the loss of revenue obtained by our model is -0.04. This is because our framework focuses on the performance of the whole system rather than approximate the value of  $\theta_1$ . The loss of revenue acts as one of the constraints. Table 5 presents a more detailed result of a 10-fold cross validation when we set  $\theta_1$  as -0.05 in Yahoo (I) dataset. The average loss of revenue is 3.6%, while the average improvement of ad CTR is 13.1% and average improvement of visual saliency is 8.0%. The performance of our proposed framework is very similar in all the folds, except fold 6. These consistent results confirm our cross validation. We observe that, in table 3, the weigh of memorability is 0.00, thus the change of memorability is quite small. This is because the optimal weights  $\gamma_k, k = 1 \dots 6$  change with threshold values  $\theta_k, k = 1 \dots 6$ . If the publisher values more about the memorability and assigns  $\theta_3$  to a relatively larger value, the corresponding weight and variable change will increase.

## 7 CONCLUSION

In this paper, we discuss a computational framework that brings multimedia metrics to RTB to optimize trade-offs among stakeholders in display advertising. We consider the contextual relevance to ensure the user's online experience and increase CTR. We consider the visual saliency and image memorability to increase the user's engagement towards displayed ads. Our experimental results show that the proposed framework is able to increase the benefits of selected advertisers and the user with just a slight decrease in the publisher's revenue. In the long run, better engagements of advertisers and users will increase the demand of advertising and supply of webpage visits, which will boost the publisher's revenue. How to effectively model the changes and engagement process in the online advertising ecosystem in a long term will be our future work. And also, analyzing the properties of the proposed framework from the aspect of game theory will be another interesting topic.

## 8 ACKNOWLEDGEMENT

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research

Centre in Singapore Funding Initiative. The authors also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

## REFERENCES

- [1] O. Ben-Zwi, M. Henzinger, and V. Loitzenbauer. Ad exchange: Envy-free auctions with mediators. In *The 11th International Conference on Web and Internet Economics*, pages 104–117, 2015.
- [2] A. Broder, P. Ciccolo, M. Fontoura, E. Gabrilovich, V. Josifovski, and L. Riedel. Search advertising using web relevance feedback. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pages 1013–1022, 2008.
- [3] R. Burke and T. Srull. Competitive interference and consumer memory for advertising. *Journal of Consumer Research*, 15(1):55–68, 1988.
- [4] D. Campbell and R. Wright. Shut-up I don't care: understanding the role of relevance and interactivity on customer attitudes toward repetitive online advertising. *Journal of Electronic Commerce Research*, 9(1):62, 2008.
- [5] P. Chatterjee, D. Hoffman, and T. Novak. Modeling the clickstream: implications for web-based advertising efforts. *Marketing Science*, 22(4):520–541, 2003.
- [6] X. Chen, B. Chen, and M. Kankanhalli. Optimizing trade-offs among stakeholders in real-time bidding by incorporating multimedia metrics. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017.
- [7] X. Chen, T. Nguyen, and M. Kankanhalli. SalAd: a multimodal approach for contextual video advertising. In *2015 IEEE International Symposium on Multimedia*, pages 211–216, 2015.
- [8] Hubspot. Horrifying display advertising. <https://blog.hubspot.com/marketing/horrifying-display-advertising-stats>, 2015.
- [9] IAB. Iab tech lab content taxonomy. <https://www.iab.com/guidelines/iab-quality-assurance-guidelines-qag-taxonomy/>, 2016.
- [10] A. Khosla, A. Raju, A. Torralba, and A. Oliva. Understanding and predicting image memorability at a large scale. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2390–2398, 2015.
- [11] A. Khosla, J. Xiao, P. Isola, A. Torralba, and A. Oliva. Image memorability and visual inception. In *SIGGRAPH Asia 2012 Technical Briefs*, page 35. ACM, 2012.
- [12] X. Lu, X. Zhao, and L. Xue. Is combining contextual and behavioral targeting strategies effective in online advertising? *ACM Transactions on Management Information Systems*, 7(1):1, 2016.
- [13] D. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [14] S. McCoy, A. Everard, P. Polak, and D. Galletta. The effects of online advertising. *Communications of the ACM*, 50(3):84–88, 2007.
- [15] M. Neumeier. *The Brand Gap: How to Bridge the Distance Between Business Strategy and Design*. New Riders, 2005.
- [16] J. Owens, B. Chaparro, and E. Palmer. Text advertising blindness: the new banner blindness? *Journal of Usability Studies*, 6(3):172–197, 2011.
- [17] P. Sajjacholapunt and L. Ball. The influence of banner advertisements on attention and memory: human faces with averted gaze can enhance advertising effectiveness. *Frontiers in Psychology*, 5, 2014.
- [18] F. Sarić, G. Glavaš, M. Karan, J. Šnajder, and B. Bašić. Takelab: systems for measuring semantic text similarity. In *Proceedings of the 1st Joint Conference on Lexical and Computational Semantics*, pages 441–448, 2012.
- [19] Y. Yuan, F. Wang, J. Li, and R. Qin. A survey on real time bidding advertising. In *IEEE International Conference on Service Operations and Logistics, and Informatics*, pages 418–423, 2014.
- [20] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech. Minimum barrier salient object detection at 80 fps. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1404–1412, 2015.